Method for the expression of unknown environmental DNA into adapted host cells

This application is the US national phase of international application PCT/EP2003/007765 filed 17 July 2003 which designated the U.S. and claims benefit of EP 02291871.8, dated 24 July 2002, the entire contents of each of which are hereby incorporated by reference.

Introduction and Background

The present invention relates to methods and compositions for nucleic acid production, analysis and cloning. The present invention discloses tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify novel pathways, novel enzymes and novel metabolites of interest in various areas, including pharmaceutical, cosmetic, agrochemical and/or food industry.

Drug discovery process is based on two main fields, namely combinatorial chemistry and natural products. Combinatorial chemistry has shown its ability to generate huge amounts of molecules, but with limited chemical diversity. At the opposite, natural products have been the most predominant source of structural and molecular diversity. However, the exploitation of this diversity is strongly hampered by their limited access, complex identification and purification processes, as well as by their production.

Microorganisms are known to synthesize a large diversity of natural compounds which are already widely used in therapeutic, agriculture, food and industrial areas. However, this promising approach to the identification of new natural compounds has always been considerably limited by the principal technological bolts of isolating and *in vitro* propagating the huge diversity of bacteria. Most microorganisms living in a natural, complex environment (soil, digestive tract, sea, etc…) have not been cultivated because their optimal living conditions are either unknown or difficult to reproduce. Numbers of scientific publications relate this fact and it is now assumed that less than about 1% of

the total bacterial diversity (when all environments are considered together) have been isolated and cultivated (Amann *et al*, 1995).

New approaches have been developed to try to overpass the critical step of isolation, and to access directly to the huge genetic potential established by the microbial adaptation processes through their long evolution. These approaches are called "Metagenomic" because they address a plurality of genomes of a whole bacterial community, without any distinction (metagenome).

Metagenomics involve direct extraction of DNAs from environmental samples and their propagation and expression into a cultivated host cell, typically a bacteria. Metagenomic has been firstly developed for the identification of new bacterial phylum (Pace .1997). This use is based on the specific cloning of genes recognized for their interest as phylogenetic markers, such as 16S rDNA genes. Further developments of Metagenomics relate to the detection and cloning of genes coding for proteins with environmental or industrial interest. These first two applications of metagenomic involve a first step of gene selection (generally using PCR) before cloning. In the case of protein production, the cloning vector used are preferentially also expression vectors, i.e., they contain regulatory sequences upstream of the cloning site causing expression of the cloned gene in a given bacterial host strain.

More recent developments of metagenomic consider the total metagenome cloned without any selection and/or identification, to establish random "Metagenomic DNA libraries". This provides an access to the whole genetic potential of bacterial diversity without any "*a priori*" selection. Metagenomic DNA libraries are composed of hundreds of thousands of clones which differ from each other by the environmental DNA fragments which have been cloned. In this respect, large DNA fragments have been cloned (more than 30 Kb), so as to (i) limit the number of clones which have to be analysed and (ii) to be able to recover whole biosynthetic pathways for the identification of new metabolites resulting from multi enzymatic synthesis. This last point is of particular interest for bacterial metagenomic libraries since, most the biosynthetic pathways have been found to be naturally organised in a same cluster of DNA and even in the same operon in bacteria. Nevertheless, the heterologuous expression of a whole

biosynthetic pathways (large DNA fragment) needs a much more improved system than a simple expression vector to have a full and stable expression.

Except for the identification and characterisation of bacterial community at the phylogenetic or diversity levels, metagenomic libraries produced in the prior art are gene expression libraries, i.e., the environmental DNA fragments are cloned downstream of a functional promoter, to allow their expression and analysis. In this regard, WO99/45154 and WO96/34112 relate to combinatorial gene expression libraries which comprise a pool of expression constructs where each expression construct contains DNA which is operably associated with one or more regulatory regions that drive expression of genes in an appropriate organisms. Furthermore, the expression constructs used in these methods have a very limited and invariable host range. Similarly, WO 01/40497 relates to the construction and use of expression vectors which can be transferred in one chosen expression bacterial host of the *Streptomyces* genus. All these approaches are, however, very limited since they require the presence of expression signals and confer invariable or very limited host range capabilities. Furthermore, most (if not all) metagenomic DNA libraries have been established in *E. coli* which is the most efficient cloning system. However, most environmental DNA are not expressed or functionally active in *E. coli*. In particular, functional analysis in *E. coli* of genes cloned from G+C rich organisms, such as Actinomyces, could be limited by the lack of adequate transcription and translation system. Also, posttranslational modification system in *E. coli* is not operative on heterologous proteins from Actinomicetes and some specific substrates for proteins activity are not present in *E. coli*.

The stable maintenance of large foreign DNA fragments (> 10Kb) into a selected host cell is one of the key points for academic research or applied industrial purposes. Usually, the vector carrying the foreign DNA is maintained by cultivating the host cells in a medium with a vector-specific selective pressure (resistance to an antibiotic for example). However, when large foreign DNA fragments are cloned and/or expressed, their propagation and/or expression require energy, which is not allocated for cell growth anymore. As a consequence of this new resource allocation (nutrients/energy), it is not unusual to have a genetic rearrangement of the foreign DNA (deletion,

modification etc…) as a recombinant cell reaction. This results in the modification of the foreign genetic information and in the loss of DNA functionality. This can be observed without any loss of the selective pressure carried by the vector. As a result, the recombinant clone is no more exploitable for genetic or functional analysis.

Thus, the exploitation of the huge potential of metagenomics for the discovery of new natural compounds, pathways or genes cannot be achieved with currently existing methods. Alternative technologies and processes must be developed, to allow stable maintenance and propagation of large foreign DNAs into host cells for production of efficient libraries and functional screening in a large variety of host cell species, including *Bacillus* or *Streptomyces*, to take full account of the huge diversity of the environmental DNAs.

Summary of the Invention

The present invention discloses improved tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify and produce novel pathways, novel enzymes and novel metabolites of interest.

More particularly, the invention now proposes to keep the advantage of high efficient cloning in *E. coli* and to modify the properties of metagenomic libraries, to allow genetic and functional analyses of particular selected clones in any appropriate system, thereby making possible the stable maintenance and propagation, the analysis and/or the expression of the huge diversity of metagenomic libraries. According to the invention, polynucleotide libraries can be produced in any convenient cloning system, such as *E. coli*, and then modified, depending on the desired selection or screening system, to adapt host range and/or properties of the library (or a portion thereof).

A particular object of this invention resides more specifically in a method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the method comprising (i) selecting cloning vectors in the

library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors into a selected host cell and integration of the polynucleotide contained in said vectors into the genome of the selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.

An other object of this invention is a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing (i) transfer of the environmental DNA into a selected host cell distinct from *E. coli*, (ii) integration of the environmental DNA into the genome of a selected host cell, and (iii) stable maintenance and propagation of the environmental DNA into the selected host cell.

A further object of this invention is a method of producing modified libraries of polynucleotides, the method comprising selecting a sub-population of clones in a first library, based on the presence or absence of a characteristic of interest, and modifying the properties of said selected clones to allow their functional analysis or expression.

The modification in the library or cloning vector is typically obtained by targeted insertion of a polynucleotide construct, preferably using transposable elements, either in vitro or in vivo.

The integration into the genome of the selected host cell is typically obtained by site specific integration or by homologous or heterologous DNA/DNA recombination.

The invention is particularly suited for producing and analysing genetic diversity (metagenomic libraries), to identify new genes and isolate new metabolites, drugs, enzymes, antibiotics, etc.

<u>Legend to the Figures</u>

Figure 1 : Map of pP1 vector carrying transposable construct Tn<Apra> fig.1a. Excised transposable construct is 1132 bps in size. It contains two mosaic ends (ME) and a gene conferring resistance to apramycine (Apra) fig.1b.

Figure 2 : pPL1 Vector (fig. 2a) that carries the conjugative transposable construct Tn<Apra-oriT> (fig. 2b). The nucleotide sequence of the transposable construct contains an origin of transfer (oriT) and a gene conferring resistance to apramycine (Apra). Direction of DNA transfer at oriT is shown by an arrow.

Figure 3 : Transposable construct Tn<Apra-oriT-att-int> (fig. 3a) on vector pPAOI6 (fig. 3b). The transposable construct contains ΦC31 integrase gene and attachment DNA sequence for site specific integration, origin of transfer and gene for selection. The orientation of genes and direction of DNA transfer are marked by arrows.

Figure 4 : This figure shows any target DNA suitable for insertion of transposable construct (fig. 4a). Insertion of conjugative and site specific integrative transposable construct Tn<Apra-oriT-int> is shown on fig. 4 b and c. Insertion of transposable construct into selective gene marker carried on original vector is shown on fig. 4b. In this event, the cloned insert is intact and can be transferred to heterologous host. Insertion of transposable construct into cloned DNA insert results in gene inactivation (fig. 4c).

Figure 5 : Morphological differences between *Streptomyces* transconjugant (assay). Conjugations have been performed with FS3-124 modified with transposable construct pPAOI6; (control) conjugation have been performed with pPAOI6.

Figure 6: Schematic Map of pPSB vector (top) and transposable elements (bottom). The transposable element has 720 bp DNA of the *amyE* gene from *B. subtilis* in addition to *att-int*-oriT-*Apra*.

Figure 7: Plasmid pPSBery (top) and transposable element (bottom) carrying selective marker *ery* AM, for resistance to erithromycine, and part of *amy E* gene for homologus recombination in *B. subtillis*.

Figure 8: Integrase Φ C31 was deleted from pPSBery plasmid. Resulting plasmid is pPSBery-DI (top). Transposable element contains *Apra* and *ery* AM genes for selection, oriT origin of transfer and a part of *amyE* gene for integration in to *amyE* locus of *B. subtilis* chromosome (bottom).

Figure 9: Map of pTn5-7 AOI plasmid. Transposable element has ends of tn5 (ME) and tn7 (T7 R, T7 L) transposons.

Detailed Description of the Invention

The invention provides novel strategies, methods and products for generating and analysing combinatorial gene libraries. As indicated above, the invention discloses, particularly, methods of analysing libraries of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the methods comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors and/or expression of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell, such as by genetic, biochemical, chemical or phenotypical approaches.

In a most preferred embodiment, the methods allow stable transfer and propagation of large environmental nucleic acids in a selected host following initial selection. Such methods comprise (i) selecting cloning vectors in a library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors and integration of the polynucleotide which they contain into a selected host cell genome, and (iii) analysing the polynucleotides

contained in said modified vectors upon transfer of said modified vectors into said selected host cell, such as by genetic, biochemical, chemical or phenotypical approaches.

Library of polynucleotides

The term “library of polynucleotides” designates a complex composition comprising a plurality of polynucleotides, of various origins and structure. Typically, the library comprises a plurality of unknown polynucleotides, i.e., of polynucleotides whose sequence and/or source and/or activity is not known or characterized. In addition to such unknown (or uncharacterized) polynucleotides, the library may further include known sequences or polynucleotides. Typically, the library comprises more than 20 distinct polynucleotides, more preferably at least 50, typically at least 100, 500 or 1000. The complexity of the libraries may vary. In particular, libraries may contain more than 5000, 10 000 or 100 000 polynucleotides, of various origin, source, size, etc. Furthermore, the polynucleotides are generally cloned into cloning vectors, allowing their maintenance and propagation in suitable host cells, typically in *E. coli*. The polynucleotides in the library may be in the form of a mixture or separated from each other, in all or in part. It should be understood that some or each polynucleotide in the library may be present in various copy numbers.

The polynucleotides in the libraries are more preferably obtained or cloned from complex sources of nucleic acids, most preferably from environmental samples. Such libraries are also termed “metagenomic libraries” since they contain nucleic acids derived from whole genomes of mixed populations of microorganisms.

The term environmental sample designates, broadly, any sample containing (a plurality of) uncharacterized (micro)organisms, particularly uncultivated (or non-cultivable) microorganisms. The sample may be obtained or derived from specific organisms, natural environments or from artificial or specifically created environments (e.g., industrial effluents, etc). An uncultivated (or non-cultivable) microorganism is a microorganism that has not been purposely cultured and expanded in isolated form. The

sample may be obtained or derived from soil, water, mud, vegetal extract, wood, biological material, marine or estuarine sediment, industrial effluents, gas, mineral extracts, sand, natural excrements, meteorits etc. The sample may be collected from various regions or conditions, such as tropical regions, deserts, volcanic regions, forests, farms, industrial areas, household, etc.

Environmental samples usually contain various species of (uncharacterized, uncultivated) microorganisms, such as terrestrial microorganisms, marine microorganisms, salt water microorganisms, freshwater microorganisms, etc. Species of such environmental microorganisms include autotrophe or heterotrophe organisms, eubacteria, archaebacteria, algae, protozoa, fungi, viruses, phages, parasites, etc. The microorganisms may include extremophile organisms, such as thermophiles, psychrophiles, psychrotophes, acidophiles, halophiles, etc. More specific examples of environmental bacteria includes actinomycetes, eubacteriaes and mycobacteriaes, examples of fungi include phycomycetes, ascomycetes and basidiomycetes, etc. Other organisms include yeasts (saccharomyces, kluyveromyces, etc.) plant cells (algae, lichens, etc), corals, etc. for instance. The sample may comprise various species of such donor (uncultivated) microorganisms, as well as various amounts thereof. The environmental sample may contain, in addition, known and/or cultivable microorganisms (e.g., prokaryotic or eukaryotic), as well as nucleic acids and organic materials. The sample may also contain different animal cells: mammalian cells; insect cells, etc (arising from larvae, feces, etc.).

It should be understood that the present invention is not limited to any specific type of sample or environmental microorganism, but can be used to produce diversity, create nucleic acid libraries, etc., from any environmental sample comprising uncultivated microorganisms. The sample may be wet, soluble, dry, in the form of a suspension, paste, powder, solid, etc. Preferably, the sample is dry or in solid or semi-solid state (e.g., paste, powder, mud, gel, etc.). The sample may be treated prior to nucleic acid extraction, for instance by washings, filtrating, centrifuging, diluting, drying, etc.

The term “environmental DNA” designates any DNA fragment or collection obtained from an environmental sample. Nucleic acids may be extracted / isolated from the sample according to various techniques, such as those described in WO01/81357, in WO01/40497, in Handelsman *et al*. (Chemistry & Biology 5(10), 1998, R245), Rondon *et al*. (Tibtech 17, 1999, 403 ; Applied and Environm. Microbiol. 66, 2000, 2541), Miller et al (Applied and Environm. Microbiol. 65, 1999, 4715) or Frostegard *et al*. (Applied and Environm. Microbiol., 65, 1999, 5409).

In a particular embodiment of the above method, the library comprises a plurality of environmental DNA fragments. The library may also comprise other types of nucleic acids, such as environmental RNAs, for instance.

The polynucleotides have a size which is typically comprised between 10 and 100 kb, more preferably between 20 and 80 kb, typically between 30 and 80 kb. Although not mandatory, it is preferred that the polynucleotide fragments in the library all have similar size, to produce homogenous libraries.

Cloning Vectors

As indicated, the polynucleotides are contained or cloned into cloning vectors. These vectors may be of various types, including plasmids, cosmids, fosmids, episomes, artificial chromosomes, phages, viral vectors, etc. In most preferred embodiments, the cloning vectors are selected from plasmids, cosmids, phages (e.g., P1-derivatives) and BACs, even more preferably from cosmids, P1 derivatives and BACs. By using cosmids or P1 derivatives, it is possible to generate homogenous libraries, since these vectors essentially accommodate polynucleotides having a size of approximately 40 kb and 80 Kb, respectively. Furthermore, since the invention provides that the vectors are modified after the initial cloning step, the cloning capacity of the vectors is maximized and inserts of as much as 40 and 80 kbs in length can be cloned into fosmids, BAC and P1 derivatives.

As indicated the cloning vector has a particular host range, i.e., the ability to replicate in a particular type of host cell. Typically, the host is a bacteria, more preferably an *E. coli* strain. Indeed, *E. coli* is so far the most convenient host cell for performing recombinant technologies. The advantage of the present invention is that the starting library can be produced in any suitable host system of choice, since the properties of the libraries will be adapted later during the process.

Cloning vectors generally comprise the polynucleotide insert and genetic elements necessary and sufficient for maintenance into a competent host cell. They typically contain, in addition to the polynucleotide insert, an origin of replication functional in a selected host cell as well as a marker gene for selection and screening. The cloning vector may comprise additional elements, such as promoter regions, for instance. Although cloning vectors may replicate in several different host cells, they are usually adapted to a particular host cell type and not suitable or efficient for replication or maintenance in other cell types.

In a preferred embodiment, the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids , BAC or P1 derived vectors. *E. coli* cloning vectors may carry an origin of replication derived from naturally-occurring plasmids, such as ColE1, pACYC and p15A, for instance. Many *E.coli* cloning vectors are commercially available and/or can be constructed using available regulatory sequences.

Screening of the cloning vectors

In step i) of the method, a first selection or screen is performed on the polynucleotide library. The screen is performed so as to identify or select clones having (or lacking) a particular, common characteristic. The selection may be carried out according to various techniques, such as molecular screening, protein expression, functional screening, etc. A preferred selection is performed by molecular screening. Molecular screening designates any method of identification of molecular or structural characteristics in a polynucleotide sequence. This can be made by a variety of techniques which are known per se, such as hybridisation, amplification, sequencing, etc. Preferably, molecular

screening comprises the selection of clones in a library which contain, in their sequence, a particular sequence or region or motif, said sequence or region or motif being characteristic of a particular type of activity or gene (enzyme, biosynthetic pathways, etc.).

In a first variant, the selection is made by contacting the cloning vectors in the library with a particular nucleic acid probe (or set of probes) containing a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.). The cloning vectors in the library which hybridise to the probe (or set of probes) are then selected.

In a second variant, the selection is made by contacting the cloning vectors in the library with a particular pair of nucleic acid primers specific for a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.), and a PCR amplification reaction is performed. The cloning vectors in the library which lead to a positive amplification product are then selected.

In this regard, the present application provides new primers designed in conserved motives of the β-keto acyl synthase gene, which are particularly useful for screening polynucleotides containing putative polyketide synthase (PKS) genes or domains. These primers have the following degenerated sequence :

Sense primer : 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3' (SEQ ID NO: 3)

Antisense primer : 5'- GCBBSSRYYTCDATSGGRTCSCC -3' (SEQ ID NO: 4)

wherein :

R is A or G

S is G or C

Y is C or T

K is G or T

D is A or G or T, and

B is C or G or T.

A particular object of this invention is a polynucleotide primer having one of the above sequences, typically a mixture of different polynucleotide primers having a sequence corresponding to one of the above degenerated sequences. A particular object of this invention also resides in a pair of primers each having one of the above sequences.

Once particular clones have been selected, the analysis of their polynucleotides needs to be confirmed and/or validated, and/or their polynucleotides can be used to study their function and/or produce novel compounds or metabolites.

The invention now enables such further analysis and uses, by allowing a modification of the cloning vectors that is specific and adaptable by the skilled person, depending on the activity which is sought. In particular, it is possible to confer properties such as specific expression or a novel, specific host range to the selected cloning vectors, to assess their activity, as disclosed below.

<u>Modification of the cloning vectors</u>

After high efficiency cloning using most convenient cloning vectors such as BACs or cosmids propagated into *E. coli*, and after the identification, selection and/or characterisation of cloned DNA fragments, the invention now allows to modify specifically the cloning vectors to transfer, integrate into the genome, maintain, express and/or over-express the selected polynucleotides into any selected host expression system, which is suitable to assess the selected activity or property. Such selected hosts may be native or heterologous host cells, and include, but are not limited to, for example *Streptomyces*, *Nocardia*, *Bacillus*, fungi, yeasts, etc.

The selected cloning vectors of the library may be modified according to various techniques. The modification is typically a genetic modification, comprising the introduction of particular genetic sequences into the structure of the cloning vector, in addition to or in replacement of sequences contained in said vector. It is highly preferred to use specific or targeted (or oriented) techniques to improve the efficacy of the method. By “specific” is meant that the modification occurs at a pre-determined location

in the cloning vector, through site-specific mechanisms. By "targeted" is meant that the modification occurs in a controlled way, so as not to alter the polynucleotide insert contained in the vector in a non-desirable way.

In a preferred embodiment, the selected vectors are modified by insertion, into the vector, of a target polynucleotide construct which contains genetic elements conferring the selected property(ies) to the cloning vector.

The target polynucleotide construct typically comprises the genetic elements necessary to transfer, propagate, integrate into the genome, maintain, express or overexpress the cloned polynucleotide into a chosen (bacterial) host expression system. Said genetic elements may include particular origin(s) of replication, particular origin(s) of transfer, particular integrase(s), transcriptional promoter(s) or silencer(s), either alone or in combination(s).

In a first, preferred variant, the target polynucleotide construct comprises a genetic element allowing transfer of the vector into a selected host cell.

Natural DNA transfer mechanisms between donor and recipient strains is known under the term conjugation or conjugative transfer. Conjugative transfer can occur between different strains of the same species as well as between strains of different species. Many naturally occurring plasmids carry so called *tra* genes, which are involved in and mediate conjugative transfer. The DNA transfer starts at specific DNA structures, known as an origin of transfer or "ori T". The presence of such an oriT in a vector allows said vector to be transferred into a desired host cell.

In a particular, preferred embodiment, the target polynucleotide construct comprises an origin of transfer functional in the selected host cell.
The structure of various oriT has been reported in the art (Guiney *et al*. 1983; Zechner *et al*. 2000). In a specific embodiment, the origin of transfer is selected (or derived) from RP4, pTiC58, F, RSF1010, ColE1 and R6K(α).

A specific example of an oriT which can be used in the present invention derives from plasmid RP4 and has or comprises all or a functional part of the following sequence (SEQ ID NO: 5) :

gatctGTGATGTACTTCACCAGCTCCGCGAAGTCGCTCTTCTTGATTGGAGCGCATGGG
GACGTGCTTGGCAATCACGCGCACCCCCCGGCCGTTTTAGCGGCTAAAAAAGTCATG
GCTCTGCCCTCGGGCGGACCACGCCCATCATGACCTTGCCAAGCTCGTCCTGCTTCT
CTTCGATCTTCGCCAGCAGGGCGAGGATCGTGGCATCACCGAACCGCGCCGTGCGC
GGGTCGTCGGTGAGCCAGAGTTTCAGCAGGCCGCCCAGGCGGCCCAGGTCGCCATT
GATGCGGGCCAGCTCGCGGACGTGCTCATAGTCCACGACGCCCGTGATTTTGTAGCC
CTGGCCGACGGCCAGCAGGTAGGCCGACAGGCTCATGCCGGCCGCCGCCGCCTTTT
CCTCAATCGCTCTTCGTTCGTCTGGAAGGCAGTACACCTTGATAGGTGGGCTGCCCT
TCCTGGTTGGCTTGGTTTCATCAGCCATCCGCTTGCCCTCATCTGTTACGCCGGCGGT
AGCCGGCCAGCCTCGCAGAGCAGGATTCCCGTTGAGCACCGCCAGGTGCGAATAAG
GGACAGTGAAGAAGGAACACCCGCTCGCGGGTGGGCCTACTTCACCTATCCTGCCC
GGCTGACGCCGTTGGATACACCAAGGAAAGTCTACACGAACCCTTTGGCAAAATCC
TGTATATCGTGCGAAAAAGGATGGATATACCGAAAAAATCGCTATAATGACCCCGA
AGCAGGGTTATGCAGCGGAAAAGATCCGTCGGATCT

The term “functional part” designates any fragment or variants of the above sequence which retain the capacity to cause conjugative transfer. Such fragments typically comprise at least 80%, preferably at least 85% or 90% of the above sequence. Variants may include one or several mutations, substitutions, deletions or additions of one or several bases.

In an other particular variant, the target polynucleotide construct comprises a genetic element allowing integration of the vector (or of the polynucleotide contained therein) into the genome of the selected host cell.

A donor DNA is permanently or stably maintained and expressed in a selected recipient cell if it is integrated into the recipient cell’s genome or if it contains elements that allow autonomous replication in said cell. In a most preferred embodiment, the vector is modified to allow transfer and integration of the polynucleotide into the host cell genome.

Integration is a preferred way of ensuring stable expression. Integration can be obtained by physical recombination. Recombination can be homologous, e.g., between two homologous DNA sequences, or illegitimate, where recombination occurs between two

non-homologous DNAs. As a particular example, integration of donor DNA into the chromosome of the recipient can be mediated by host recombination repair system or by site-specific recombination. Another well-studied process that can transfer and integrate genes is transduction by bacterial viruses, such as λ and ϕC31. In a phage–infected bacterial cell, fragments of the host DNA are occasionally packaged into phage particles and can then be transferred to a recipient cell. Integration into the recipient cell's genome is caused by an integrase.

In a specific embodiment, the target polynucleotide construct comprises a nucleic acid encoding an integrase functional in the selected host cell. More preferably, the integrase is selected from λ and ϕC31 integrases. In a specific embodiment, the polynucleotide construct comprises a nucleic acid encoding an integrase having or comprising all or a functional part of the following sequence of the ϕC31 integrase (SEQ ID NO: 6) :

AGATCTCCCGTACTGACGGACACACCGAAGCCCCGGCGGCAACCCTCAGCGGATGC
CCCGGGGCTTCACGTTTTCCCAGGTCAGAAGCGGTTTTCGGGAGTAGTGCCCCAACT
GGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAGGGTCGCCGACATGACACAAG
GGGTTGTGACCGGGGTGGACACGTACGCGGGTGCTTACGACCGTCAGTCGCGCGAG
CGCGAGAATTCGAGCGCAGCAAGCCCAGCGACACAGCGTAGCGCCAACGAAGACA
AGGCGGCCGACCTTCAGCGCGAAGTCGAGCGCGACGGGGGCCGGTTCAGGTTCGTC
GGGCATTTCAGCGAAGCGCCGGGCACGTCGGCGTTCGGGACGGCGGAGCGCCCGGA
GTTCGAACGCATCCTGAACGAATGCCGCGCCGGGCGGCTCAACATGATCATTGTCTA
TGACGTGTCGCGCTTCTCGCGCCTGAAGGTCATGGACGCGATTCCGATTGTCTCGGA
ATTGCTCGCCCTGGGCGTGACGATTGTTTCCACTCAGGAAGGCGTCTTCCGGCAGGG
AAACGTCATGGACCTGATTCACCTGATTATGCGGCTCGACGCGTCGCACAAAGAATC
TTCGCTGAAGTCGGCGAAGATTCTCGACACGAAGAACCTTCAGCGCGAATTGGGCG
GGTACGTCGGCGGGAAGGCGCCTTACGGCTTCGAGCTTGTTTCGGAGACGAAGGAG
ATCACGCGCAACGGCCGAATGGTCAATGTCGTCATCAACAAGCTTGCGCACTCGAC
CACTCCCCTTACCGGACCCTTCGAGTTCGAGCCCGACGTAATCCGGTGGTGGTGGCG
TGAGATCAAGACGCACAAACACCTTCCCTTCAAGCCGGGCAGTCAAGCCGCCATTC
ACCCGGGCAGCATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCCGTGCCGACC
CGGGGCGAGACGATTGGGAAGAAGACCGCTTCAAGCGCCTGGGACCCGGCAACCGT
TATGCGAATCCTTCGGGACCCGCGTATTGCGGGCTTCGCCGCTGAGGTGATCTACAA
GAAGAAGCCGGACGGCACGCCGACCACGAAGATTGAGGGTTACCGCATTCAGCGCG
ACCCGATCACGCTCCGGCCGGTCGAGCTTGATTGCGGACCGATCATCGAGCCCGCTG
AGTGGTATGAGCTTCAGGCGTGGTTGGACGGCAGGGGGCGCGGCAAGGGGCTTTCC
CGGGGGCAAGCCATTCTGTCCGCCATGGACAAGCTGTACTGCGAGTGTGGCGCCGT
CATGACTTCGAAGCGCGGGGAAGAATCGATCAAGGACTCTTACCGCTGCCGTCGCC
GGAAGGTGGTCGACCCGTCCGCACCTGGGCAGCACGAAGGCACGTGCAACGTCAGC
ATGGCGGCACTCGACAAGTTCGTTGCGGAACGCATCTTCAACAAGATCAGGCACGC
CGAAGGCGACGAAGAGACGTTGGCGCTTCTGTGGGAAGCCGCCCGACGCTTCGGCA
AGCTCACTGAGGCGCCTGAGAAGAGCGGCGAACGGGCGAACCTTGTTGCGGAGCGC
GCCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGGCAGGCGCGTA

CGACGGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACAGGCAGCGCTGACGCTCC
GGCAGCAAGGGGCGGAAGAGCGGCTTGCCGAACTTGAAGCCGCCGAAGCCCCGAA
GCTTCCCCTTGACCAATGGTTCCCCGAAGACGCCGACGCTGACCCGACCGGCCCTAA
GTCGTGGTGGGGGCGCGCGTCAGTAGACGACAAGCGCGTGTTCGTCGGGCTCTTCGT
AGACAAGATCGTTGTCACGAAGTCGACTACGGGCAGGGGGCAGGGAACGCCCATCG
AGAAGCGCGCTTCGATCACGTGGGCGAAGCCGCCGACCGACGACGACGAAGACGAC
GCCCAGGACGGCACGGAAGACGTAGCGGCGTAGCGAGACACCCG

The term “functional part” designates any fragment or variants of the above sequence which retain the capacity to cause integration. Such fragments typically comprise at least 80%, preferably at least 85% or 90% of the above sequence. Variants may include one or several mutations, substitutions, deletions or additions of one or several bases.

In a more preferred variant, the target polynucleotide construct comprises genetic elements allowing transfer of the cloning vector into the selected host cell and integration of the cloning vector or a portion thereof into the genome of the selected host cell. Most preferred polynucleotide constructs comprises an oriT and a nucleic acid encoding an integrase.

In an other variant, the target polynucleotide construct comprises an origin of replication specific for or functional in the selected host cell. The origin of replication may be selected (or derived), for instance, from pAMβ1, pSa, 2μm circle, pSam2, pSG1, pIJ101, SCP2, pA387 and artificial chromosomes.

In an other variant, the target polynucleotide construct comprises a transcriptional promoter functional in the selected host cell. As indicated above, in a particular variant, the invention allows to modify the cloning vector to enable expression or over-expression of the cloned polynucleotides in the selected host. The expression of genes is driven mainly by transcriptional promoters, which initiate gene transcription. The type of promoter to be used in the present invention can be selected by the skilled person, depending on the selected host cell and type of expression needed. Promoters may be ubiquitous or cell-specific, regulated or constitutive, weak or strong. They may be of various origins, including promoters isolated from viruses, phages, plant cells, bacterial

genes, mammalian genes, etc., or they may be artificial or chimeric. Typical examples of promoters include T7, T4, LacZ, trp, ara, SV40, tac, λPL, GAL, AOX, hsp-70, etc.

The target polynucleotide construct is typically a DNA molecule, although RNAs may also be used as starting material. It is typically a double-stranded DNA. The target polynucleotide construct may be produced by conventional recombinant DNA techniques, including DNA synthesis, cloning, ligation, restriction digestion, etc. and a combination thereof.

The target polynucleotide construct is preferably engineered so as to be inserted in a region of the vector distinct from the polynucleotide. Indeed, it is important that the integrity of the polynucleotides is preserved. Directed insertion may be accomplished in a variety of ways, including site-specific insertion using particular enzymatic systems (Cre/Lox, FLP, etc.), homologous recombination with particular target sequences present in the vector, or by the use of transposons or transposable elements and appropriate selection means.

In a particular, preferred embodiment, the target polynucleotide construct is contained in or comprises a transposable nucleic acid construct. Indeed, in a preferred variant, the methods of the present invention use transposable elements to alter the properties of the cloning vectors, and allow their transfer, maintenance, expression or over-expression in a selected host cell.

Transposable nucleic acid constructs are derived from transposons, which are genetic elements capable of moving from one genetic loci to another. Two main classes of transposons have been identified in bacteria. The most simple transposons comprise an insertion sequence that carries only elements of transposition. These elements are two inverted DNA repeats and a gene that codes for a protein called transposase. The transposase catalyses the excision and integration of the transposon. It has been shown that the excision and integration reaction can be catalysed in trans by a transposase, which can be provided in vivo or in vitro in purified form or expressed from a different

construct. More complex transposons carry more insertion sequences and additional genes that are not involved in transposition.

Transposable nucleic acid constructs of this invention thus typically comprise, flanked by two inverted repeats, the target polynucleotide construct and, more preferably, a marker gene. In the presence of a transposase, these transposable nucleic acid constructs can integrate into a cloning vector in vivo or in vitro, thereby providing for targeted polynucleotide insertion. Alternatively, such nucleic acid constructs can be used for targeted integration, in the absence of a transposase, in particular strains such as hypermutator strains. Such transposable nucleic acid constructs also represent a particular object of the present application. In this regard, in a more preferred embodiment, the invention also relates to a transposable nucleic acid construct, wherein said construct comprises an origin of transfer flanked by two inverted repeats. Specific examples of such construct are transposons pPL1 and pPAOI6, as disclosed in the experimental section. The transposable nucleic acid construct may further comprise an integrase gene and/or a marker gene.

The inverted repeat nucleic acid sequences may be derived from the sequence of various transposons, or artificially created. In particular, transposable elements can be generated using inverted repeats obtained from transposons or transposable elements such as Tn5, Tn21, miniTn5, T7, T10, Tn917, miniTn400, etc. Preferably, the sequences derive from transposon Tn5. In a specific embodiment, they comprise all or a functional part of the following sequences :

- left arm of pPAOI6 transposon (SEQ ID NO: 7)
CTGTCTCTTATACACATCTCAACCATCATCGATGAATTTTCTCGGGTGTTCTCGCATA
TTGGCTCGAATTCGAGCTCGGTACCC

- right arm of transposon pPAOI6 (SEQ ID NO: 8)
GATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGCCAACGACTACGCACTAGCC
AACAAGAGCTTCAGGGTTGAGATGTGTATAAGAGACAG

The marker gene may be any nucleic acid encoding a molecule whose presence in a cell can be detected or visualized. Typical marker genes encode proteins conferring

resistance to antibiotics, such as apramycine, chloramphenicol, ampiciline, kanamycine, spectinomycine, thiostrepton, etc. Other types of markers confer auxotrophy or produce a label (e.g., galactosidase, GFP, luciferase, etc).

In a specific embodiment, the cloning vector in the library comprises a first marker gene and the modification step ii) comprises:

. contacting in vitro, in the presence of a transposase, the selected cloning vectors with a transposon comprising, flanked by two inverted repeats, the target polynucleotide construct and a second marker gene distinct from the first marker gene, and

. selecting the cloning vectors which have acquired the second marker gene and which have lost the first marker gene.

The double selection ensures that the target polynucleotide construct has been inserted at a site within the marker gene present in the cloning vector, i.e., outside of the polynucleotide insert.

It should be understood that the modification may be accomplished in various other ways, particularly by incorporating a sequence coding for the transposase directly into the transposable element or into another expression unit. The presence and expression of the transposase can be regulated by inductive promoter or termosensitive replicative units. Also, transposition can be carried out by *in vitro* process.

<u>Analysis of the polynucleotides</u>

In step (iii) of the process, the polynucleotides may be analysed by various methods, including by genetic, biochemical, chemical or phenotypical approaches, which are well known per se in the art. Analysis occurs upon transfer and, optionally, expression of the polynucleotides into the selected host cell.

In this regard, the modified cloning vectors can be transferred into the selected host cell by a variety of techniques known in the art, including by transformation, electroporation, transfection, protoplast fusion, conjugative transfer, etc. In a preferred embodiment, the

target polynucleotide construct comprises an oriT and the modified vectors are transferred into the selected host cells by conjugative transfer. In this embodiment, the cloning vector and the selected host cells are co-cultivated and the recombinant host cells are selected and isolated.

The selected host cell may be any type of cell or microorganism, including, without limitation, *Steptomyces, E. coli, Salmonella, Bacillus, Yeast, fungi*, etc.

One of the objectives of the invention is to be able to analyse environmental DNAs of unknown cellular origin into different host expression systems. In order to analyse the potentiality of the DNAs at the transcription and/or translation levels and to have much more probabilities to have a DNA expression, it is important to have the possibility to test different host expression systems.

Insertion of foreign DNA into host expression systems like *Streptomyces* can produce, for instance, an increase in doubling time, morphological modifications, pigments production, etc., which can be related either directly to the expression of foreign DNA or by combinatorial biology of the foreign DNA and the biology of the expression host systems. The new phenotypes can be analysed by all techniques known in the art such genetic, biochemical, chemical, phenotypic approaches etc.

In a preferred, specific embodiment, the invention relates to a method for the identification or cloning of polynucleotides encoding a selected phenotype, the method comprising (i) cloning environmental DNA fragments into *E. coli* cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer and integration in a selected host cell, (iv) transferring the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

By applying the above method, new polynucleotide sequences have been identified, cloned and characterized, which produce new phenotypes in bacteria. These polynucleotides contain the sequence of PKS genes and other genes that encode polypeptides involved in biosynthetic pathways. The sequence of these polynucleotides is provided in SEQ ID NO:1 and SEQ ID NO:2.

The complete annotated DNA sequences of fosmid clones FS3-124 and FS3-135 are provided in SEQ ID NOs: 1 and 2, respectively, and as further described in the following CDS information.

Specifically, the following CDS information are related to SEQ ID NO:1.

```
CDS             76..1134
                /note="ABC_transportr"
                /gene="TAP2 PROTEIN"
                /blastp_match="Oryzias latipes"
                /blast_score= 0.002
CDS             complement(1096..2430)
                /note="none"
                /blastp_match="Anabaena sp"
                /gene="ALR1117 PROTEIN"
                /blast_score=2e-18
CDS             1178..1624
                /note="Gram_pos_anchor"
                /gene="CELL WALL SURFACE ANCHOR"
                /blast_score=1e-04
                /blastp_match="Streptococcus pneumoniae"
CDS             complement(2506..3567)
                /note="CONSERVED"
                /gene="HYPOTHETICAL PROTEIN"
                /blast_score=0.019
                /blastp_match="Deinococcus radiodurans"
CDS             complement(2906..4222)
                /note="glycosyl transferase"
                /gene="lipopolysaccharide"
                /blast_score=2e-23
CDS             complement(4092..5321)
                /note="glycosyl transferase"
                /gene="glycosyl transferase"
                /blast_score=1e-15
CDS             complement(6337..8502)
                /note="PUTATIVE"
                /gene="GLUTAMINE AMIDOTRANSFERASE"
                /blastp_match=" Bordetella bronchiseptica"
                /blast_score=1e-16
CDS             complement(8181..9530)
                /note="none"
                /gene="MEMBRANE PROTEIN"
                /blast_score=0.035
CDS             complement(9531..10721)
                /note="NOEC Transmembrane"
                /gene="NODULATION PROTEIN"
                /blastp_match="Azorhizobium caulinodans"
                /blast_score=3e-07
CDS             complement(10504..11274)
                /note="PUTATIVE"
                /gene="HYDROLASE"
                /blastp_match="Streptomyces coelicolor"
                /blast_score=4e-14
CDS             12874..13689
                /note="HYPOTHETICAL Meth-transf"
```

```
                /gene="PROTEIN PA1088"
                /blast_score=2e-06
                /blastp_match=" Pseudomonas aeruginosa"
CDS             14195..15976
                /note="PUTATIVE, Glyco_transf"
                /note=""
                /gene="LIPOPOLYSACCHARIDE BIOSYNTHESIS"
                /blast_score=2e-06
                /blastp_match="Vibrio cholerae"
CDS             15427..16512
                /note="PATHWAY: INNER CORE LIPOPOLYSACCHARIDE
                BIOSYNTHESIS"
                /gene="PHOSPHOHEPTOSE ISOMERASE "
                /blast_score=3e-17
                /blastp_match="Helicobacter pylori"
CDS             15579..16253
                /note="none"
                /gene="PHOSPHOHEPTOSE ISOMERASE"
                /blast_score=2e-22
                /blastp_match="Neisseria meningitidis"
CDS             complement(16505..17656)
                /note="BIOSYNTHESIS PUTATIVE"
                /gene="LIPOPOLYSACCHARIDE"
                /blast_score=2e-17
                /blastp_match="Thermotoga maritima"
                /pfam_match="Glycos_transf_1"
CDS             complement(17657..18697)
                /note="none"
                /gene="ALR3073 PROTEIN"
                /blast_score=6e-27
                /blastp_match="Anabaena sp"
CDS             complement(18615..19304)
                /note="none"
                /gene="ALR4487 PROTEIN"
                /blast_score=8e-07
                /blastp_match="Anabaena sp"
CDS             complement(19301..20596)
                /note="ATP_GTP_A"
                /gene="ABC TRANSPORTER"
                /blast_score=3e-61
                /blastp_match="Synechocystis sp"
CDS             complement(20535..21476)
                /note="PERMEASE COMPONENT"
                /gene="POLYSACCHARIDE ABC TRANSPORTER"
                /blast_score=6e-41
                /blastp_match="Clostridium acetobutylicum"
CDS             complement(22025..22951)
                /note="involved in the synthesis of a polysaccharide
                capsule ?"
                /gene="32.3 KDA PROTEIN"
                /blast_score=3e-17
                /blastp_match="Sphingomonas sp"
CDS             23155..26523
                /note="peptide syntase"
                /gene="mcyA, mcyB and mcyC"
                /blastp_match="Microcystis aeruginosa"
                /blast_score=0.0
CDS             26409..34433
                /note="polyketide syntase et peptide syntase"
                /gene="mcyD, mcyE, mcyF and mcyG"
                /blastp_match="Microcystis aeruginosa"
CDS             34418..37500
                /note="CYSTATIN"
                /gene="PEPTIDE SYNTHETASE"
                /blast_score=0.0
                /blastp_match="Anabaena sp"
CDS             35359..37500
                /note="gene cluster"
                /gene="nostopeptolide biosynthetic"
                /blast_score=2e-41
```

```
                /blastp_match="Nostoc sp"
   Sequence 37500 BP; 6199 A; 12698 C; 12769 G; 5834 T; 0 other;
```

The following CDS information are related to SEQ ID NO:2.

```
   CDS               complement(3..914)
                     /blast_score=2e-66
                     /blastp_match="AE004644.PA2177 Pseudomonas aeruginosa"
                     /gene="regulator hybrid"
                     /note="probable similarity to prokaryote sensory
                     transduction proteins"
                     /product="sensor/response regulator hybrid"
   CDS               924..2168
                     /note="none"
                     /gene="ligase"
                     /blastp_match="AP003013.MLR8297 Mesorhizobium loti"
                     /blast_score=e-152
                     /product="2-amino-3-ketobutyrate CoA ligase"
   CDS               2207..3190
                     /blast_score=e-151
                     /blastp_match="AE008872.TDH  Salmonella typhimurium"
                     /note="none"
                     /gene="dehydrogenase"
                     /product="threonine 3-dehydrogenase"
                     /pfam_match="PF00107; adh_zinc; 1"
   CDS               3373..4455
                     /note="putative"
                     /gene="methyltransferase"
                     /blastp_match="AE001866.DR0026 Deinococcus
radiodurans"
                     /blast_score=1e-08
   CDS               4546..4959
                     /blast_score=2e-11
                     /blastp_match="AP002997.MLL1617 Mesorhizobium loti"
                     /gene="unknown"
                     /note="pfam00263, GSPII_III, Bacterial type II and III
                     secretion system protein, Expect = 7.8"
   CDS               5176..6192
                     /blast_score=5e-98
                     /blastp_match="AF064070.PE20 Burkholderia
pseudomallei"
                     /gene="glucose epimerase"
                     /note="putative "
                     /product="UDP-glucose 4-epimerase"
                     /pfam_match="PF01370; Epimerase; 1"
   CDS               6331..14043
                     /note="substrat AT, malonyl ; zinc depend
dehydrogenase ;
                     zinc dependent adenosine deaminase putative"
                     /gene="PKS I"
                     /blastp_match="AF285636.WCBR  2547 Burkholderia
mallei"
                     /blast_score=0.0
                     /pfam_match="ketoacyl-synt"
                     /pfam_match="ketoacyl-synt_C"
                     /pfam_match="Acyl_transf"
                     /pfam_match="SAM binding"
                     /pfam_match="adh_zinc"
```

```
                     /pfam_match="pp-binding"
   CDS               14275..15408
                     /blast_score=e-104
                     /blastp_match="AF285636.WCBT Burkholderia mallei"
                     /gene="acyl-CoA transferase WcbT"
                     /note="putative"
                     /pfam_match="PF00155; aminotran_1_2; 1"
   CDS               15436..16245
                     /blast_score=5e-11
                     /blastp_match="TTDEFFMT.FMT T.thermophilus"
                     /gene="formyltransferase"
                     /note="evidence experimental"
                     /product="methionyl-tRNA formyltransferase"
                     /pfam_match="PF00551; formyl_transf; 1"
                     /pfam_match="PF02911; formyl_transf_C; 1"
   CDS               16287..17384
                     /note="putative"
                     /gene="glycotransferase"
                     /blastp_match="AF285636.WCBD Burkholderia mallei"
                     /blast_score=e-99
   CDS               17427..18158
                     /blast_score=2e-82
                     /blastp_match="AF285636.WZM Burkholderia mallei"
                     /gene="ABC-2 transporter Wzm"
                     /note="putative"
                     /pfam_match="PF01061; ABC2_membrane; 1"
   CDS               18248..18847
                     /blast_score=7e-61
                     /blastp_match="AF285636.WZT Burkholderia mallei"
                     /gene="ABC-2 transporter Wzt"
                     /note="putative "
                     /pfam_match="PF00005; ABC_tran; 1"
   CDS               18952..20346
                     /note="putative"
                     /gene="glycosyltranferase"
                     /blast_score=e-101
                     /blastp_match="AF285636.WCBE Burkholderia mallei"
                     /pfam_match="Glycos_transf_1"
   CDS               20442..21167
                     /note="putative"
                     /gene="unknow"
                     /blast_score=1e-26
                     /blastp_match="AE009248.ATU3189 Agrobacterium
tumefaciens"
   CDS               complement(21164..24301)
                     /note="tranporter domain"
                     /gene="unknow"
                     /blast_score=5e-29
                     /blastp_match="AE009122.ATU1658 Agrobacterium
tumefaciens"
                     /prosite_match="PS00402; BPD_TRANSP_INN_MEMBR;
UNKNOWN_1"
   CDS               complement(24351..27023)
                     /note="none"
                     /gene="unknow"
                     /blast_score=5e-29
                     /blastp_match="AP003581.ALR0267 Nostoc sp"
   CDS               complement(27806..29686)
                     /note="none"
                     /gene="cell surfarce protein"
                     /blast_score=2e-34
                     /blastp_match="AE010748.MA0851  2567 Methanosarcina
```

```
                     acetivorans"
                     /product="cell surface protein"
   CDS               complement(29535..30872)
                     /note="none"
                     /gene="cell surface protein"
                     /blast_score=2e-36
                     /blastp_match="AE010748.MA0851 Methanosarcina
acetivorans"
   CDS               complement(30848..32647)
                     /note="fragment"
                     /gene="O-antigen"
                     /blast_score=1e-72
                     /blastp_match="AF105060.RFBC Riftia pachyptila"
                     /product="O-antigen biosynthesis protein"
                     /pfam_match="PF00535; Glycos_transf_2"
   CDS               complement(32574..35555)
                     /note="putative"
                     /gene="glycosyltransferase"
                     /blast_score=7e-46
                     /blastp_match="AE013462.MM2213 Methanosarcina mazei "
   CDS               complement(35533..36598)
                     /blast_score=7e-37
                     /blastp_match="AE013462.MM2213 Methanosarcina mazei"
                     /gene="glycosyltransferase"
                     /note="putative"
   CDS               complement(36516..37400)
                     /blast_score=8e-22
                     /blastp_match="AP003581.ALR0267 Nostoc sp"
                     /gene="ALR0267"
                     /note="putative ATP/GTP-binding protein, esterase fush
                     9e-14"
   Sequence 37507 BP; 5531 A; 13507 C; 13011 G; 5458 T; 0 other;
```

The invention also relates to any polynucleotide sequence comprising all or part of these sequences (i.e., SEQ ID NOs: 1 or 2), their complementary strand, or a functional variant thereof. A part of the above sequences includes, preferably, at least 20 consecutive bases, more preferably at least 50 consecutive bases thereof, even more preferably a coding sequence (e.g., an CDS). In this respect, SEQ ID NOs: 1 and 2 comprise several novel open reading frames encoding novel polypeptides involved in biosynthetic pathways. These coding sequences are identified above.

In a specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides (CDS) 76 - 1134 ; 1096 - 2430 ; 1178 - 1624 ; 2506 - 3567 ; 2906 - 4222 ; 4092 - 5321 ; 6337 - 8502 ; 8181 - 9530 ; 9531 - 10721 ; 10504 - 11274 ; 12874 - 13689 ; 14195 - 15976 ; 15427 - 16512 ; 15579 - 16253 ; 16505 - 17656 ; 17657 - 18697 ; 18615 - 19304 ; 19301 - 20596 ; 20535 - 21476 ; 22025 -

22951 ; 23155 - 26523 ; 26409 - 34433 ; 34418 - 37500 and 35359 - 37500 of SEQ ID NO: 1 or a complementary strand thereof.

In an other specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides 3 - 914, 924 - 2168 ; 2207 - 3190 ; 3373 - 4455 ; 4546 - 4959 ; 5176 - 6192 ; 6331 - 14043 ; 14275 - 15408 ; 15436 - 16245 ; 16287 - 17384 ; 17427 - 18158 ; 18248 - 18847 ; 18952 - 20346 ; 20442 - 21167 ; 21164 - 24301 ; 24351 - 27023 ; 27806 - 29686 ; 29535 - 30872 ; 30848 - 32647 ; 32574 - 35555 ; 35533 - 36598 and 36516 - 37400 of SEQ ID NO: 2 or a complementary strand thereof.

Variants of these sequences include any naturally-occurring variant comprising or or several nucleotide substitutions ; sequences variants resulting from the degeneracy of the genetic code, as well as synthetic variants coding for functional polypeptides. Variants include any sequence that hybridise under high stringent conditions, as disclosed for instance in Sambrook et al., to any of the above sequences, and encode a functional polypeptide. The invention also include any nucleic acid molecule encoding a polypeptide comprising all or a fragment of an amino acid sequence encoded by a polynucleotide as disclosed above. Preferably, the fragment comprises at least 10 consecutive amino acid residues, more preferably at least 20, even more preferably at least 30.

The invention also relates to any vector comprising these polynucleotide sequences. These sequences may be DNA or RNA, preferably DNA, even more preferably double-stranded DNA. The invention also relates to a polypeptide encoded by a polynucleotide sequence as defined above. The invention also relates to a method of producing such polypeptides by recombinant techniques, comprising expressing a polynucleotide as defined above in any suitable host cell and recovering the encoded polypeptide. The invention also relates to a recombinant host cell comprising a polynucleotide or a vector as defined above.

An other object of this invention resides in a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing transfer and integration of the environmental DNA into the genome of a selected host cell distinct from *E. coli*.

The sub-DNA-libraries should have either desired genetic characteristics based on high or low GC content, DNA encoded for a desired enzymatic activity, part or full biosynthetic pathways for metabolites etc., or specific origin such as soil fractions, animal organs, sub fraction of a microorganism community etc. The invention also allows to produce conjugative vector with desired characteristics in accordance with the characteristics of the pre-identified sub-DNA-libraries and functional analysis of mutants in heterologous hosts. It can also be used, without limitation, for the production of mutants by mutagenesis, for DNA sequencing, genes or biosynthetic pathways knock-out by insertion or to confer transfer capabilities for expression, co-expression, over-expression or modification of biosynthetic pathways.

Further aspects and advantages of the invention will be disclosed in the following examples, which should be regarded as illustrative and not limiting the scope of this application.

Experimental Section

A - From *E.coli* to *Streptomyces*

In this work, we constructed a fosmid library in *E. coli* from total DNA prepared directly from soil. The library has been screened for presence of biosynthetic pathways. We developed genetic tools for functional genomics that allow gene identification, inactivation and horizontal gene transfer from *E. coli* to *Streptomyces*.

The cloning vectors in the library contain ColE1 replicon for propagation in *E. coli*. Transposable elements based on Tn 5 transposon were produced and used for *in vitro* modification of selected cloning vectors. Integrated transposable elements contain gene for resistance to apramicyne. Conjugative derivatives were constructed by incorporating origin of transfer from plasmid RP4. A conjugative and site specific integrative transposon was also constructed comprising the integrase gene from ϕ C31 phage, including attP attachement site. Conjugal transfer was demonstrated from an appropriate *E. coli* donor cell to another *E. coli* or *Streptomyces lividans* recipient cell. Constructed transposon was tested for inactivation of the genes cloned into fosmids. Obtained mutants can be used for direct sequencing by adequate primers and transferred by conjugation into *Streptomyces lividans*. Transposable elements thus represent very useful tools for functional analysis of a large DNA libraries cloned into BAC, PAC fosmids or other cloning vectors in which cloned inserts must be transferred into heterologous host.

Materials and methods

Bacterial strains, plasmids and growth conditions.

E. coli DH10B (F- mcrA delta(mrr-hsdRMS-mcrBC) phi80dlacZ deltaM15 delta lacX74 deoR recA1 endA1 araD139 delta (ara, leu)7697 galU galK lambda- rpsL nupG), strain (Epicentre) was used for fosmid and plasmid transformation and DNA amplification.

Unless specifically described, all DNA manipulations were performed according to Sambrook, J., *et al*, (1989).

Soil DNA extraction and DNA libraries construction.

Total bacterial community DNA was extracted and large DNA libraries have been constructed into fosmids according to the method described in WO 01/81357.

Fosmid DNA extraction and Purification

Fosmids DNA containing soil librairy were extracted from pools of 96 clones. Culture of recombinant clones were performed in Deep-Well 96 and 48, respectively in 1 ml and 2 ml of LB media containing 12,5 $\mu g\ ml^{-1}$ of chloramphenicol. Cultures were grown at

37°C with shaking at 250 RPM during 22 hours. DNA extraction was done by using the Nucleobond PC100 extraction kit (Macherey Nagel).

PCR screening for the detection of PKS genes

Primers design

Degenerate PCR primers sets were designed to specifically amplify PKS nucleic acids sequences. Multiple sequence alignment of PKS domains revealed highly conserved motives, in particular in the β-keto acyl synthase domain. Primers Lib1F and Lib2R were designed in conserved motives of the β-keto acyl synthase gene. Lib1F (sense primer, 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3') and Lib2R (antisense primer, 5'- GCBBSSRYYTCDATSGGRTCSCC -3') were deduced from β-keto acyl synthase peptide sequences GP(AS)(LV)(AST)(IV)DTAC and GDPIE(TVA)(RAQ)A, respectively. The specific fragment amplified with Lib1F / Lib2R was approximately about 465 bp (corresponding to 155 amino acids). Specificity and efficiency of the PCR systems were validated by testing on positive DNA controls (i.e. genomic DNA from type I PKS producing strain such as *Bacillus subtilis*, *Streptomyces lividans, Streptomyces ambofaciens* and *Ralstonia solanocearum*) and negative DNA controls (genomic DNA from strains which are known to do not contain PKS genes). Furthermore, DNA extracted from soil samples were tested to calibrate PCR techniques.

PCR conditions

PCR conditions were optimised, in particular for concentrations of DMSO, $MgCl_2$, Primers and DNA template quantities. For PCR using microorganism genomic DNA and soil DNA as template (50 to 200 ng), the PCR mix (50µl) contained 250µM of dNTP, 5 mM $MgCl_2$ final, 2,5% DMSO, 1X PCR buffer, 0,75 µM of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For PCR using fosmid pooled DNA as template (100 to 500 ng), the PCR mix (50µl) contained 250µM of dNTP, 5 mM $MgCl_2$ final, 5% DMSO, 1X PCR buffer, 0,75 µM of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For identification of positives clones in 96 microtiter plates, 25 µl of each bacteria culture were used as template and PCR conditions were the same as above. Thermocycling program was : a denaturation

step at 96°C for 5 minutes; then 1 minute at 96°C, 65°C for 1 minute, 72°C for 1 minute. The first 7 cycles, the annealing temperature was lowered 1°C per cycle until 58°C was reached. A subsequent 40 cycles were carried out with the annealing temperature at 58°C. A final extension step was at 72°C for 7 minutes. For identification of positives clones in 96 microtiter plates, the first denaturation step of 96°C was during 8 minutes. The other steps were the same as described above. PCR reactions were performed with a PTC 200 thermocycler (MJ Research).

PCR products analysis

PCR products of about 465 bp were purified on agarose gel with gel extraction Kit (Qiagen) according to the manufacturer recommendations. First approach consisted in subcloning PCR products using the Topo PCR II kit (Invitrogen). Recombinant Plasmids were extracted using QIAprep plasmid extraction Kit (Qiagen) and sequenced with Forward and Reverse M13 primers with CEQ 2000 automated sequencer (Beckman Coulter). Second approach consisted in direct sequencing of PCR products. Sequencing data were compared with nucleic and proteic genbank database using BLAST program.

Sequencing of the identified fosmid insert DNA and sequence analysis

Fosmids inserts were sequenced using either a transposon-mediated and by shotgun subcloning approach. Transposition was realized by using (Transposition Kit) commercialized by Epicentre according to the manufacturer. For shotgun subcloning, transformants were grown for 16 hours at 37°C. Fosmid extraction was done by using the Nucleobond PC100 extraction kit (Macherey Nagel). DNA was partially restricted with *Sau3A* and sized on standard gel elecrtphoresis for fragments ranging from 1 to 3 Kbs and cloned into Bluescript vector according to Sambrook *et al*. (1989).
Sequence analysis was performed with the identification of ORFs by using Frameplot of the GC3. Each identified ORF was compared to gene databases by using BLAST program. PKS domains were determined by aligning obtained sequence versus already described PKS domains from domain databases.

Sequencing

Sequencing reactions were performed with 1 μg of DNA and 3.6 pmol of primer, using CEQ 2000 Dye Terminator Cycle Sequencing kit (Beckman Coulter) under conditions proposed by supplier. Ten μL of reaction products were precipitated using 4 μL of solution containing 1.5 M NaOAc, 50 mM EDTA and 60 μL cold 95% ethanol/dH_2O from –20 °C. The pellet was washed 2 times with 200 μL 70% ethanol/dH_20, vacuum dried and dissolved in 40 μL sample loading solution (supplied in kit). Sequencing reactions were run on an CEQ 2000 sequencer (Beckman Coulter).

Plasmids construction and validation

Plasmid pP1 was constructed as follows. A 941 bp DNA fragment containing native promoter region and AA(3)IV gene was amplified by polymerase chain reaction (PCR) using primers AmF (d-CCCTAAGATCTGGTTCATGTGCAGCTCCATC, SEQ ID NO: 9) and AmR (d-TAGTACCCGGGGATCCAACGTCATCTCGTTCTCC, SEQ ID NO: 10). One hundred microlitre reaction were performed containing 0.1μM each of primers, 1 X Vent DNA polymerase buffer (NEB), 0.2 μM of each deoxyribonucleoside triphosphate (dNTP), 50 ng of the DNA template and 2U of Vent DNA polymerase (NEB). PCR mixture was heated for 4 min at 94 °C in a PTC-200 thermocycler (Peltier) and cycled 25 x at 94 °C for 60 sec, 59 °C for 30 sec and at 72 °C for 70 sec. The final extension was performed at 72 °C for 7 min.

PCR product was purified using GFX DNA purification kit (Amersham), then digested by *Bgl* II and *Sma* I restriction enzymes. A 941 bp BglII/SmaI fragment was inserted into the *Bam* HI, *Sma* I sites of pMOD plasmid (EPICENTRE). DH10B *E. coli* was transformed with pP1 and subjected to apramicyne selection on LB agar plates. Six colonies surviving on apramicyne selection were grown in liquid LB media and final pPL1 candidates were thoroughly checked via PCR and restriction mapping.

To construct conjugative plasmid pPL1, 750 bp oriT DNA region from plasmid RP4 were amplified via PCR using primers oriTF (d-GCGGTAGATCTGTGATGTACTTCACCAGCTCC, SEQ ID NO: 11) and oriTR

(TAGTACCCGGGGATCCGACGGATCTTTTCCGCTGCAT, SEQ ID NO: 12). PCR conditions were as above. Amplified DNA was digested using *Bgl* II and *Sma*I restriction enzymes. *Bgl*II/*Sma*I DNA fragment was subjected to purifiation after gel electophoresis on 0.7% agarose. Purified fragment was ligated into pP1 plasmid digested by *Bam* HI and *Sma* I restriction enzymes.

ϕ C31 integrase gene and attachment site (attP) was amplified via PCR using primers Fint (d-AACAAAGATCTCCCGTACTGACGGACACACCG, SEQ ID NO: 13) and PJ (d-CGGGTGTCTCGCATCGCCGCT, SEQ ID NO: 14). Amplified DNA fragment was purified by GFX kit (Amersham) and phosphorilated using T4 polynucleotide kinase (NEB) under conditions recommended by the enzyme manufacturer. Phosphorilated DNA fragment was cloned in to pPL1 vector opened with SmaI restriction enzyme (NEB) and dephosphorilated by calf alkaline phosphatase (NEB). DH10B *E coli* was transformed with ligation mixture using Bio Rad Pulsing apparatus and protocols provided by Bio-Rad. Twelve transformants were analyzed by PCR for the presence of integrase gene. Orientation of integrase gene was verified by restriction analysis using *Bgl* II and *Eco*RI restriction enzymes. Resulting plasmid was named pPAOI6.

To construct pPAOI6-A plasmid, pPAOI6 plasmid was digested with *Eco*RI and *Bgl*II restriction enzymes followed bydigestion by Bean mung nuclease (NEB). Linearised plasmid was self ligated and transformed in to DH10B cells.

Plasmid preparation of fosmid DNA

Fosmid and BAC DNA for sequencing was prepared by using the Nucleobond AX kit (Macherey-Nagel), following protocol for BACs, Cosmid as specified by manufacturer.

Mutagenesis

Transposon Tn-pPAOI6 was prepared by digestion of pPAOI6 plasmid using *Pvu*II restriction enzymes, followed separation on agarose gel and purification of fragment containing transposon from gel using Qiagen kit. The same molar ratio of transposon and corresponding fosmid was used for mutagenesis in vitro using Tn5 transposase (Epicentre) and conditions specified by manufacturer. We transformed aliquot of the transposed mixture by electroporation into competent DH10B *E. coli* strain.

Conjugation *E. coli- Streptomyces lividans TK24*

Conjugation experiments were done using 6 x 10^6 *E coli* S17.1 cells containing conjugative plasmids or fosmids. The *E coli* cells were grown in LB media with adequat antibiotic. The cells were collected by centrifugation, washed two times using same volume of LB media and concentrated to 10^8 cells/ml and overlaid on LB plates containing 2x 10^6 pregeminated *Streptomyces lividans* TK24 spores. The cells mixture were grown over night at 30^0 C and *E. coli* cells were washed three times using 2 ml of LB media. The plates were overlayed using top agar containing NAL (nalidixic acid) and the appropriated antibiotic. Plates were incubated for 4 days at 30°C and transformant streptomyces colonies were isolated on HT medium (Pridham *et al*. 1957) containing NAL and the same appropriated antibiotic.

Results

Construction of a Transposon Tn <Apra>.

E. coli aminoglycoside -(3)- acetyl transferase IV gene (aa(3)IV) was amplified by PCR and cloned in pMOD vector (Epicentre) . Advantage of this selective marker allow positive selection in *E. coli* and in *Streptomyces lividans*. Transposon can be used for insertional inactivation in vitro using purified transposase Tn5 (Epicentre). The structure of the pP1 constructed vector and transposon was shown on Figure 1a,1b.

Construction of a conjugative plasmid-transposon.

Conjugative vector, transposon was constructed by cloning origin of transfer from plasmid RP4 into pP1 vector producing pPL1 vector (Figure 2a, 2b). The origin of transfer was cloned in such orientation that the selective aa(3)IV gene is the last transferred during conjugation. PPL1 vector was introduced in to specific *E. coli* S17.1 strain that carry RP4 plasmid integrated in to chromosome. In conjugation experiment between donor strain S17.1 carrying pPL1 plasmid and DH10 *E coli* receptor we obtained DH10B strain carrying pPL1 plasmid. This data shows that cloned oriT fragment is functional in the pPL1 plasmid. Plasmid pPL1 can be used for DNA cloning, gene inactivation by homologous recombination. Cloned genes or part of the gene cloned could be then transferred by conjugation in to another host. Another advantage of

this vector is conjugative transposon that can be excised from vector and inserted randomly in vivo in to another DNA molecule by purified Tn 5 transposase.

Construction of a conjugative, site specific integrative plasmid- transposon for horizontal gene transfer between *E. coli* and *Streptomyces* strains.

PPL1 plasmid was used to clone an integrase gene from phage ϕC31, resulting a plasmid pPAOI6 (Figure 3a, 3b). We tested several clones for horizontal transfer between *E. coli* S17.1 strain and *Streptomyces lividans* TK24 strain. The best transfer was obtained for plasmid pPAOI6 were orientation of integrase gene is in opposite orientation to the gene for resistance to apramycine. Conjugative transfer of pPAOI6 gene in to *S. lividans* strain is confirmed by resistance to apramycine or G418. Additional confirmation of transfer was obtained using PCR method. We were able to amplify 2 kb insert using specific primers for the ϕC31 integrase gene, and no PCR amplification was obtained for control *S. lividans* TK24 strain.

pPAOI6 transposon was cloned into the *Eco*RV site of plasmid pGPS3 (New England Biolabs). This construction allows transposition not only by transposase Tn5 but also using Transposase ABC (New England Biolabs). Resulting plasmid pTn5-7AOI is shown on Figure 9.

The goal of these constructions was to produce transposons that is further used in functional analysis of the metagenomic DNA library from the soil that were constructed in a laboratory (Figures 4a, 4b, 4c).

Functional analysis of the metagenomic DNA library from soil.

The fosmids library consists of 120 512 clones, containing ~ 40 kb inserts of soil DNA. The library contains approximately 4.8 Gbps of the DNA cloned from soil. Ten percents of the library was screened by using a PCR approach for the presence of the genes involved in production of secondary metabolites (PKS). Using gene-module specific set of primers we were able to identified positive clones organized in microtiter plates (96 wells). Sequences (based on PCR products) obtained from fifteen randomly positive clones indicate that the DNA library contains very little sequence redundancy limited to one and that the sequences were found to be new and very diverse in comparison to gene databases (data not schown).

Two fosmids DNA was prepared from two positive clones (FS3-124 and FS3-135) and analyzed by sequencing. DNA analysis *in silico* shows high G+C contents of 72% and 69% respectively of the cloned inserts and presence of cluster genes that could be involved in biosynthesis of secondary metabolites. No specific phenotype was observed for the two clones in *E coli*. We employed pPAOI6 transposon mutagenesis to produce conjugative mutants. Transposon mutants FS3-124::pPAOI6 and FS3-135::pPAOI6 were isolated using apramycine as selective antibiotic. Obtained mutants were then tested on LB pates containing chloramphenicol. About 1% of the tested clones are chloramphenicol sensitive. These clones contain transposon inserted into locus encoding chloramphenicol resistance gene and not into cloned DNA insert. ApraR and ChloS transposon mutants are then used for horizontal gene transfer into *Streptomyces lividans* TK24 strain. Fosmid DNA was prepared from mutants and transformed into *E. coli* S17.1 strain. Horizontal gene transfer between *E. coli* S17.1 and *Streptomyces lividans* was done due to inserted pPAOI6 transposon. Transconjugants of the *Streptomyces lividans* were tested by PCR to confirm gene transfer and integration of the conjugative fosmid in to *S. lividans* chromosome. Both transconjugants showed an increase in doubling time, morphological modifications and pigments production in comparison to the control (Figure 5).

B - From *E.coli* to *Bacillus subtilis*

Plasmid pPSB was constructed as follows : A part of amyE gene from *B subtilis* was amplified by PCR using primers amyE-BamHI : atcgcaggatcctgaggactctcgaacccg (SEQ ID NO: 15) and amyE-EcoRI : cgactgaattcagatctagcgtgtaaattccgtctgc (SEQ ID NO: 16). DNA fragment was digested by *Eco*RI and *Bam*HI restriction enzymes and ligated into *Eco*RI, *Bgl*II site of pPAOI6 plasmid. Transposable element contains Tn<amyE-int ϕC31-oriT-apra> is shown on figure 6 with plasmid pPSB.

Plasmid pPSBery (Figure 7) was constructed by cloning erm AM gene into pPSB plasmid. A 1140 bp *Sau*3A DNA fragment containing ery AM gene with his own promoter was cloned from plasmid pMUTIN (Wagner *et al.* 1998) into the Bam HI site

of plasmid pPSB. Orientation of ery gene was confirmed by sequencing. Transposable element contains Tn < amyE-int ϕC31-ery-oriT-apra> (Figure 7).

Plasmid pPSBery-ΔI was obtained after Sma I digestion and self-ligation of core plasmid (Figure 8). In this construction, ϕC31 integrase gene was deleted from transposable element. New transposon is Tn<amyE-ery-oriT-apra> (Figure 8). All transposons could be released as linear DNA by PvuII digestion from plasmids mentioned above and transposed by transposase Tn 5 (Epicentre) in vitro.

The selection of transposed elements was done using 100 μg/ml erythromycine or 40μg/ml apramycine in *E coli* or 0.3 μg/ml erythromycine in *B subtilis*. DNA was transformed by electro-transformation into electrocompetent *E coli* strains or by competence into *B. subtilis*. Integration of imported DNA into amy E locus of *B. subtilis* chromosome was confirmed using pPSBery-ΔI plasmid. Integration was confirmed by PCR using plasmid-specific and amyE locus-specific primers. Fifteen eryR *B. subtilis* clones were tested by PCR. All transformants showed integration at amyE locus of *B subtilis* chromosome, confirming the functionality of the method and constructs.

References

Brau, B., Pliz, U., and Pipersberg, W. 1984. Genes for gentamicin-(3)-N-acetyltransferases III and IV: I. Nucleotide sequence of the AAC(3)-IV gene and possible involvement of an IS 140 element in its expression. Mol Gen Genet. 193 (1), 179-187.

Guiney, D.G., Yakobson, E., (1983). Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2. Proc Natl Acad Sci USA, 80: 3595-8.

T. Kaeberlein, K. Lewis, Epstein S. S. 2002. Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. Science 296, 1127-1129.

Mazodier, P., Thompson, C., 1989. Intergeneric conjugation between E. coli and Streptomyces species. J. Bacteriol. 171, 3583-3585.
Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. Science 276: 734-740.

Pridham, T.G., Anderson P., Foley, C., Lindenfelser, L.A., Hesseltine, C.W., and Benetdict R.C. (1957). A selection of media for maintenance and taxonomic study of *streptomyces*. Antibiotics Annual 1956-1957, 947-953.

Voeykova T, Emelyanova L, Tabakov V, Mkrtumyan N. 1998. Transfer of plasmid pTO1 from *Escherichia coli* to various representatives of the order *Actinomycetales* by intergeneric conjugation. FEMS Microbiol Lett. 162(1):47-52.

Sambrook, J., Fritsch E.F. and Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor NY.

Torsvik, V., Goksoyr, J. Daae, F. L. 1990. High diversity in DNA of soil bacteria. Appl. Environ Microbiol. 56, 782-787.

Torsvik, V., Sorheim, R. and Goksoyr, J. Total bacterial diversity in soil and sediment communities a review. J. Industr. Microbiol. 17, 170-178.

Torsvik, V., Daae, F.L, Sandaa, R.A., Overas, L. 1998. Novel techniques for analysing microbial diversity in natural and perturbed environments. J. Biotechnol. 64. 53-62.

Wagner, V., Dervyn, E. and Ehrlich S. D. 1998. A vector for systematic gene inactivationin *Bacillus subtilis*. Microbiology. 144, 3097 – 3104.

Zaehner, H. and Fiedler, F. P. 1995. Fifty years of antimicrobials: past perspectives and future trends. In The need for New antibiotics: Possible Ways Forward. Fifty –Third Symposium of the Society for General Microbiology, Bath. England, UK, (Hunter, P.A., Darby, G.K. & Russell, N.J., eds), pp. 67-84.

Zechner, E.L., de la Cruz, F., Eisenbrandt, R., Grahn, A.M., Koraimann, G., Lanka, E., Muth, G., Pansegrau, W., Thomas, C.M., Wilkins, B.M. and Zatyka, M. (2000) Conjugative-DNA transfer process. In Thomas, C.M. (ed.) The horizontal gene pool. Harwood Academic Publishers.